Multimodal (multisensory) integration, in technology

Giovanna Varni, Nicolas Castagné

Multimodal (multisensory) integration, in technology

Giovanna Varni [InfoMus Lab]
Nicolas Castagne [ACROE&INPG]

As explained in the two items multi-modality [→ Multimodality, in cognitive sciences] and [→ Multimodality, in human-computer interaction], the concept of modality has (at least) two sides, depending on the domain in which it is defined (typically cognitive sciences and human computer interaction). The current item, though written in the framework of technology and system design, uses mainly the meaning of cognitive sciences: a modality is understood as a perceptual modality, and multimodality is understood as multisensory.

In cognitive sciences, the idea that the human brain realizes an integration of the various independent perceptual modalities was very developed in the framework of the computational theory of mind [→ Computational paradigm]. Since recently, however, in the field of enactive cognitive sciences [→ Enactive cognitive sciences_ 1&2], and especially under the light of the ecological approach to perception this idea has been criticized [→ Multimodal (multisensory) integration, in cognitive sciences].

In technology, for the designers of systems that use at the same time images, sound, and gestures, ensuring a coherent perceptual experience for the user is a major aim. One can note that only a couple of technological works, for example within the Enactive Interfaces NoE, aims at approaching this question by using the recent concepts offered by Enaction. Hence, for engineers, the idea of a multimodal integration in the human brain is still very vivid. It is also, at least partially, operational. Indeed, in front of the unity of human perception, the machine offers only multiple transducers, each of which addresses a unique human sensori-motor modality. Naturally, the idea that these modalities are "integrated" by the human brain in a coherent perceptual experience appears to be helpful.

This being said, given the importance of this approach today, this item reviews shortly how the idea of multimodal integration is used in the framework of traditional human-computer interfaces (HCI), especially to help in the design of multisensory systems and interfaces [→ Interface, multimodal / multisensory] (although such an approach does not match well the Enactive approach in cognitive sciences).

In HCI and computer modeling, multi-modal integration refers to two technological research areas according to whether the focus is on designing a system to be used by humans, or on designing a system able to mimic human sensory aptitudes – especially as for the "integration" of streams of various sensory signals.

How to let a user realize multimodal integration

Sarter [Sarter, 2006] reports a set of design guidelines regarding the presentation of multisensory information. More specifically, this study focuses on the following four issues.

- selection of modalities.

This first step is very crucial because the use of multiple modalities is not always needed, but its employment is strictly correlated to a large number of factors such, for example, environmental constraints and types of tasks. Furthermore, as referred also in [Spence, 2003] "*the decision to stimulate more senses actually reflects a trade-off between the benefits of utilizing additional senses and the costs associated with dividing attention between different sensory modalities*".

- mapping of modalities to tasks and types of information.

Once those modalities have been chosen, one needs to find natural relations between them, the tasks at hand, and the informative content. By exploiting different modalities, it is possible to convey the same information, creating redundancy or different modalities for different information.

- combination, synchronization and integration of modalities.

The previous step implies considerations about spatial and temporal combination and synchronization of the sensory channels involved in the interaction.

One needs to take into account that even if signals for various modalities are presented simultaneously, this synchrony does not imply necessarily simultaneity in perception. For instance, in cases of auditory-visual interactions, it can be observed that there is a maximum effectiveness when auditory event happens before the visual event (the dimension of the time window of stimulus presentation is variable). As referred in [Oviatt, 2002] "*the empirical evidence reveals that multimodal signals often do not co-occur temporally at all during human computer or natural human communication. Therefore, multimodal system designers cannot necessarily count on conveniently overlapped signals in order to achieve successful processing in the multimodal architectures they build*".

- adaptation of multi-sensory presentation to accommodate changing task context and circumstances.

Flexibility to environmental changes and user skills is a basic requirement for any system using multiple modalities. Several methodologies and strategies can be adopted to switch between modalities.

How to model the multimodal integration process?

The other approach that uses, in Technology, the concept of multimodal integration is research on novel methodologies for building biologically inspired systems able to integrate streams of various sensory signals (coming from a camera, a microphone, etc.). In this case, the interest on the concept is shifted directly from user to machine. The main goal is to design systems reflecting as much as possible the (supposed) skill of brain in processing and merging together perceptual cues afferent by different sensory modalities. Such a goal, in fact, is not only interesting for the systems it leads to, but also because the designed models of multimodal integration are, in returns, interesting in the framework of psychology.

As cited in [Boda, 2004], there are currently two architectural metaphors helping to build such systems and interfaces performing integration, according as the instant of fusion process: early fusion and late fusion. In both cases integration is performed in one step only. Another interesting reference is [Coen, 2001]. The paper of Coen presents a possible methodology to design and build systems supporting cross-modal influence, that is "systems in which sensory information is shared across all levels of perceptual processing and not just in a final integrative stage". Classes of algorithms generally used to implement the integration step exploit, for example, neuronal networks and HMM (Hidden Markov Model).

To conclude, multimodal integration is still a very open issue, not only in neurophysiology, but also in technology, and the implementation of mechanisms of sensory fusion based on the mimesis of human and animal perceptual systems are useful to better understanding natural multisensory interactions.

References

[Boda, 2004] Boda, P.P., "Multimodal integration in a wider sense". Proceedings COLING 2004 Satellite Workshop on Robust and Adaptive Information Processing for Mobile Speech Interfaces, Geneva, Switzerlands, 2004.

[Coen, 2001] Coen, H. . "Multimodal integration-a biological view". Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI'01), pp.1417-1424. Seattle, WA, 2001.

[Oviatt, 2002] Oviatt, S.L. . "Multimodal interfaces. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications", J. Jacko and A. Sears, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, 2003, chap.14, 286-304, 2002.

[Sarter, 2006] Sarter, N.B. . "Multimodal information presentation: design guidance and research challenges". International Journal of Industrial Ergonomics, vol 36, issue 5, pp. 439-445, 2006.

[Spence, 2003] Spence, C. . "Crossmodal attention and multisensory integration: implications for multimodal interface design". Proceedings of the 5th International Conference on Multimodal interfaces, p.3-3, 2003.

Related items

Computational paradigm
Enactive cognitive sciences_ 1&2
Interface, multimodal / multisensory
Multimodal (multisensory) integration,
in cognitive sciences
Multimodality, in cognitive sciences
Multimodality, in human-computer interaction